Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 4, August 2025, pp. 2721~2734
ISSN: 2302-9285, DOI: 10.11591/eei.v14i4.8971

Prediction of stock market price for investors using machine learning approach

Omobayo Ayokunle Esan[1], Dorcas Oladayo Esan[1], Femi Abiodun Elegbeleye[2]
[1]Department of Networking and Support, Walter Sisulu University, Potsdam Campus, East London, South Africa
[2]Department of Business Applications and Development, Walter Sisulu University, Potsdam Campus, East London, South Africa

Article Info

Article history:

Received Jul 6, 2024
Revised Oct 10, 2024
Accepted Nov 28, 2024

Keywords:

Financial risk management
Investors
Machine learning
Prediction
Stock price

ABSTRACT

Stock market price prediction is a challenging task that plays a crucial role in investment decision-making and financial risk management. Traditional approaches often rely on a single machine learning (ML) algorithm for predictive modeling. In this contribution, an innovative framework that integrates logistic regression (LR) with support vector machine (SVM) to improve the accuracy and reliability of stock market price prediction. Combining the strengths of both algorithms, the proposed model harnesses the interpretability of LR and the robustness of SVM to capture complex relationships in stock market data. Experiments conducted on publicly available Yahoo Finance stock dataset and the Dhaka dataset, the results show that the proposed model yielded accuracies of 97.15% and 98.86% respectively. In comparison with other models, the proposed method outperformed the other models in terms of root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), and accuracy. The contribution and importance of leveraging hybrid modelling techniques to enhance stock market price prediction and facilitate informed investment decision-making.

Corresponding Author:

Femi Abiodun Elegbeleye
Department of Business Applications and Development, Walter Sisulu University, Potsdam Campus
East London, South Africa
Email: felegbeleye@wsu.ac.za

1. INTRODUCTION

The financial market plays a crucial role in shaping the development of global economies [1]. As the stock market gains more prominence in the economic sphere, it has garnered increasing attention from the general public [2]. A stock market or share market is the accumulation of buyers and sellers with the stocks represent ownership on businesses [3]. The economic growth and advancement in technology have led to volatility of stock market data which ruled out the capability of humans' assessments of stock prices manually [4].

One theory that explains the pricing of financial assets and the reasoning behind stock market volatility is the efficient market hypothesis [5]. The efficient market hypothesis posits that in a legally sound, well-functioning, transparent, and competitive stock market, rational investors can quickly and rationally react to all market information. As a result, stock prices will accurately, sufficiently, and promptly reflect all important facts, including a company's present and future value. An enormous amount of money is involved in stock trading, and even in a fraction time difference in price prediction can result in huge profit or loss [6].

A stock market is associated with buyers and sellers dealing stocks or shares. The stock data are a sequence of the price of a given stock equally distributed by time intervals, i.e., time-series data [7], [8]. These data are very fluctuating and equivocal which is influenced by a variational factor, including government policies, company earnings reports, influential shareholders, demand and supply, and expert speculations on current events. The stock traders mainly focused on accurate prediction of the shares to maximize profit [9]. The stock market is characterized by both uncertainty and variability, making it challenging to accurately predict market trends [10].

Stock prices can be extremely volatile, influenced by a myriad of factors such as economic data, geopolitical events, market sentiment, and even natural disasters [11]. This volatility often leads to unpredictable price movements, making reliable predictions difficult. The stock market is influenced by a vast array of variables, including economic indicators, interest rates, company performance metrics, global events, and investor sentiment. The high dimensionality of this data makes it challenging to identify which factors are most significant in predicting price movements [12]. Stock markets are often characterized by random noise fluctuations that do not follow any predictable pattern. These anomalies can be caused by irrational investor behavior, speculative trading, or unforeseen events, making it hard to distinguish between genuine signals and mere noise. Investors in the stock markets have been attempting for a long time to develop methods for predicting stock prices to make more profit [5]. However, making precise and consistently correct forecasts is complicated because of various factors, such as military war or politics, which affect investors' sentiments and thus, causing stock market movements. As a result, developing a consistent and accurate method is not easy.

Machine learning (ML) techniques have been employed in stock price prediction to improve the accuracy of predictions and alleviate these difficulties. The stock market prediction has always been appealing for the researchers. There have been numerous scientific attempts to predict the stock market effectively such as decision tree (DT)-based models and support vector machine (SVM) have been utilized in stock market forecasting but no method has been discovered accurately to predict stock price but no method has been discovered accurately to predict stock price movement.

Numerous studies have been developed in the context of studying stock market price movement variations using different techniques such as ML [13], deep learning [14], artificial neural networks (ANN) [15], and artificial intelligence (AI). these techniques are used by stock experts to assist investor to properly anticipate the situation of the company, on the basis of past and present data, and the current condition in the market. and give them a direction to make decisions. While various ML algorithms have been applied to this problem, each approach has its strengths and limitations. It is imperative to develop an intelligent model that can be used in stock market price movement prediction to enhance the predictive capability of stock price accuracy issues which can be beneficial in provision of real-life solutions to the problem of the investor's decisions.

This study utilized a cooperative model built on a ML technique which combines the logistic regression (LR) and SVM models to strengthen the results of these two individual models. LR is a linear model that is effective for binary classification tasks. It works well when the relationship between the features and the target variable is approximately linear. LR is simple to interpret and provides probabilistic outputs, which are useful in understanding the confidence of predictions while SVM, particularly when used with non-linear kernels (like the radial basis function or polynomial kernels) is used to model complex, non-linear relationships. SVM finds the optimal decision boundary (or hyperplane) that maximizes the margin between different classes, making it robust to overfitting in many cases. Hence, integrating the LR and SVM, for stock price prediction capture both linear and non-linear patterns in the data.

Drawing upon the provided background information, this study raises the following inquiry: how can the data exploratory be obtained to accurately predict stock price market to aid investor in buy and selling of stocks for proper decision making? The proposed method is developed with the consideration of the above-mentioned gaps and the associated research question, which led to the following contributions;

– New ML framework: the development of a new framework that integrates LR with SVM models to leverage the complementary strengths of both algorithms for enhanced stock market price prediction. This offers an innovative approach to maximize the stock prices performance.
– Stock optimization: a method for improving the stock prediction accuracy, thereby addressing a critical aspect of real-time stock price prediction for investor is introduced.
– Performance evaluation: detailed experimental evaluations of the proposed model on publicly available datasets to test the performance of the model with stock prices.

The following sections of this paper are arranged in the following order: section 2 provides a review of the existing related techniques. Section 3 presents a detailed explanation of the methodology; section 4 discusses various experiments and evaluations of the model. Section 5 presents the discussion. The conclusion remarks are shared in section 6.

2. RELATED TECHNIQUES

The research in [16] addressed the problem of extraction of meaningful data and uncover patterns and relationship from their databases by using linear regression and moving average are used. The results obtained show that the algorithm was able to understand the past data and does not focus on the seasonal part, thereby improved prediction accuracy. The limitation of the approach is that it is limited to a particular company stock.

The issue regarding accurate prediction of stock market returns due to stock volatility in the market was presented in [17] utilizing multiple linear regression, polynomial regression, DT regressor, and random forest regressor. The results of the proposed method shows that root mean square error (RMSE) and R-square (R^2) error yield good performance on the stock prediction. The results accuracy reduces as the dataset becomes larger.

Naeem *et al.* [18] addressed the problem of reliable prediction of stock market value is challenging due to its complexity intrinsic volatility and regular changing nature. The authors used different ML was used for comparison with the stock market value prediction to build system model.to address this issue. The used of time series forecasting with prophet and artificial recurrent neural network (RNN) long short-term memory (LSTM) yield in more accurate prediction. The intrinsic volatility complex and regular changing in nature of stock market makes reliable prediction challenging was an issue introduced in [19]. This issue was addressed using ANN with backpropagation algorithm. The experimental result shows that using prediction model give accurate predictive performance. Although, the method achieved better accuracy, the limitation is that the approach does not consider the growth or pruning method for the selection of network architecture.

The work done in [20] in addressed the problem of forecasting the stock prices is difficult using a RNN and LSTM. Experimental results obtained shows that increase in the epoch give the best training result of RMSE of 0.000983 and testing RMSE of 0.00859 was achieved. One of the limitations of the approach is that the method is highly biased. Furthermore, the problem of prediction of stock prices was addressed in [21] utilizing LSTM neural network algorithm optimized by MBGB algorithm. The result of the experiment shows that the addition of data prediction related to stock news and basic information which enhances the model stability and accuracy. The limitation of the approach is that it requires too much time for prediction. Also, research in [22] addressed the problem of stock market prices which is often full of uncertainty and affected by many factors by using SVM and ANN techniques. From the results obtained, the authors claimed that the method was able to correctly predict the stock prices. The limitation of the approach is that only sentiment data are used from various news and twitter resources no historical data are considered for prediction.

According to Chahuán-Jiménez *et al.* [23], the prediction of stock market direction since the market is intrinsic volatile across the globe is a challenging issue. The authors used ML algorithm known as multiple DT to reduce forecast error and minimize risk in investment. The results obtained shows that the method outperform other existing algorithm found in the literature. The problem of stock price prediction was addressed using auto-regressive integrated moving average (ARIMA) model [24]. The result of the implementation shows that the ARIMA model has a strong potential for short-term prediction and can compete favorably with existing techniques for stock price prediction. Furthermore, Shen and Shafiq [25] worked on short-term price trend from big data using feature engineering and deep learning-based model. The proposed solution outperforms the other ML due to comprehensive feature engineering that was used. The system achieved overall high accuracy for stock market price tend prediction. One of the limitations of this approach is that it is computationally expensive.

One can see that the existing literature regarding stock market price prediction has become apparent from the limitations. Previous research has contributed immensely but often lacked comprehensive and advanced approaches to deal with the intricacies of inaccurate prediction of stock price to assist investors in investment decision making. Through a variety of noteworthy contributions like those listed above, this research greatly strengthens solutions to these weaknesses by developing a hybrid model which integrates LR and SVM to improve stock price prediction as possible means of providing adequate information to investors on stock price to make effective investment decisions. The next section gives a detailed explanation of the proposed methodology for the prediction of stock market price.

3. METHOD

This section describes the proposed method as shown in the system architecture in Figure 1. The proposed system architecture is divided into four steps: i) data collection stage, ii) data preprocessing stage, iii) feature selection stage, and iv) prediction stage. The proposed method is further detailed in the following sections.

3.1. Data collection stage

To ensure the universality of the proposed model, the stock market dataset used in this research is obtained from the Yahoo Finance stock and Dhaka stock datasets, which as openly available [26]. This data is in comma separated values (CSV) and fed into the data pre-processing layer for further processes, as in Figure 1.

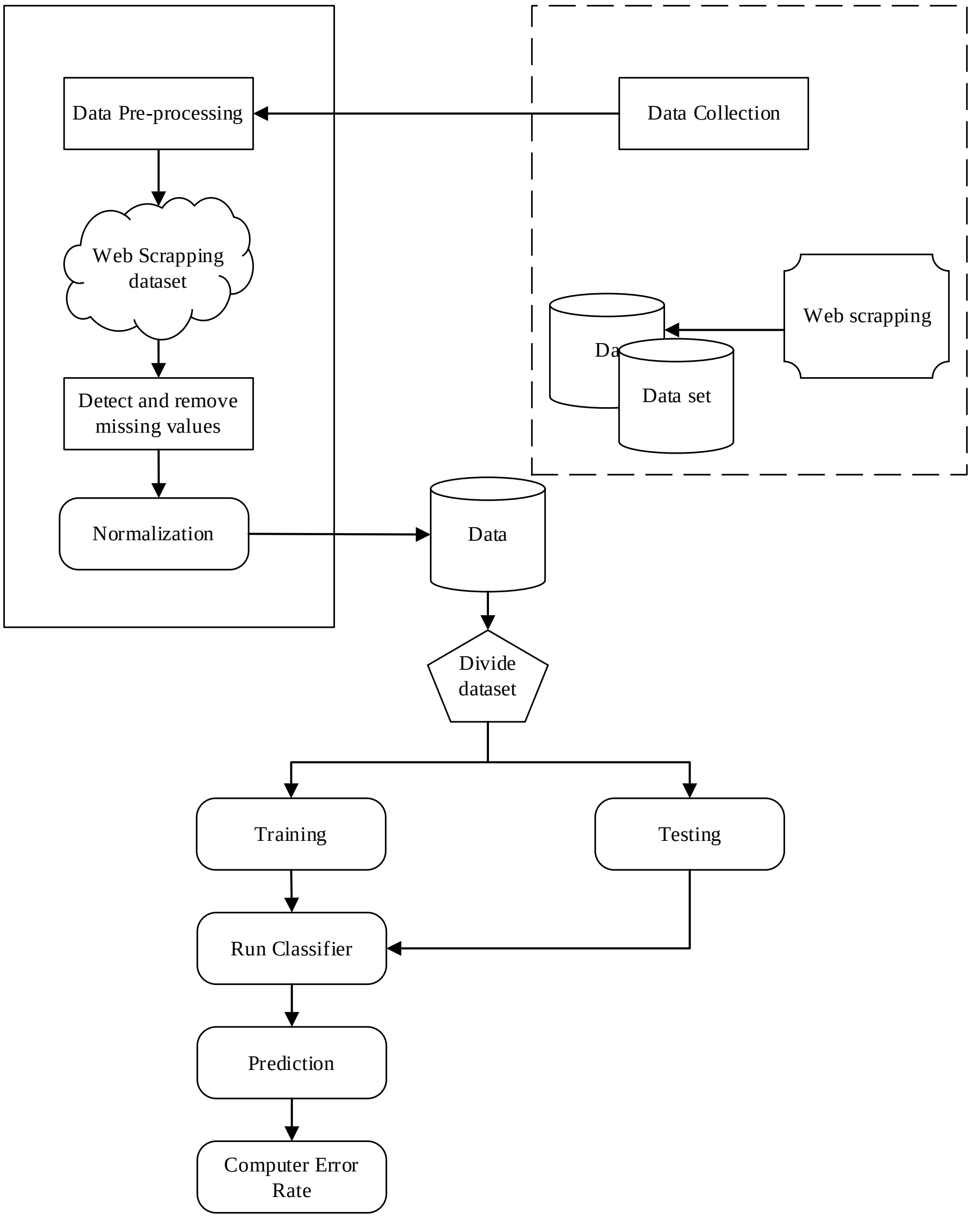


Figure 1. Proposed framework for stock price market prediction

3.2. Data pre-processing stage

At this stage, the acquired data is pre-processed to remove the outliers or noises that can affect the performance of the proposed model. In this research, the data is pre-processed through the feature scaling as explain in subsequent sub-sections.

3.2.1. Feature scaling techniques

This is a way of standardizing the independent features that are present in the data within a fixed range. The feature scaling techniques used in this study are standardization and normalization.

3.2.2. Standardization

This is a method of standardizing the independent features present in the data within a fixed. The data is scaled to a specific area to enable a thorough analysis. The variables are rescaled to have a mean of zero and the resulting distributions have a unit standard deviation. The standardized scaling as in (1):

$$X' = \frac{X-\mu}{\sigma} \tag{1}$$

where X' is the standardized data, μ is mean of the feature values, and σ is standard deviation of the feature values.

3.2.3. Normalization

This is a scaling approach in which features are re-scaled so that the data will fall in the range of zero and one. It undertakes a linear alteration on the initial data [24]. The normalized form of each feature is computed as in (2):

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{2}$$

where X_{min} is minimum value of the feature and X_{max} is maximum value of the feature.

3.3. Feature selection stage

To enhance the stock price prediction accuracy, the features from raw data using the domain knowledge. The features such as the open, high, low, close, adj_close which are extracted to improve the prediction performances of the classifiers. The output of these selected data features is passed to the prediction stage for further processing.

3.4. Prediction stage

In this study, the LR and SVM for classification of stock price. The extracted data input is passed to the LR, which is applied to the stock market data to estimate the probability of price movement direction (e.g., increase or decrease). This stage provides valuable insights into the influence of individual features on stock price changes and offers interpretability, allowing investors to understand the underlying drivers of market dynamics. The SVM is employed to refine the predictions generated by LR and capture complex nonlinear relationships that may not be fully captured by the linear model [3] using (3):

$$Y = \beta_1 X_1 + \beta_2 X_2 + \cdots \ldots . \beta_n X_n \tag{3}$$

where $X_1, X_2, \ldots . X_n$ represents the independent variables while $\beta_1, \beta_2, \ldots . \beta_n$ represent the coefficient in the LR model.

3.4.1. Support vector machine

SVM utilizes the kernel technique with the support vector to deal with the classification problems. The one-class SVM is utilized for the identification of the appropriate region that contains the unknown large data distribution. This can be represented mathematically as in (4) and (5) respectively.

$$\min_{\omega,\varepsilon,\rho} \frac{1}{2}\|\omega\|^2 + \frac{1}{vn}\sum_{i-1}^{n} \varepsilon_i - \rho \tag{4}$$

$$\text{subject to } \langle \omega_i \Phi(x_i) \rangle \geq \rho - \varepsilon_i, \varepsilon_i \geq 0 \tag{5}$$

where $x_i = (x_1, x_2 \cdots x_n)$, $i = (1, \cdots, n)$ are $n$training samples in the original data space, and ε_i is the slack variable for penalizing the outlier. The hyper-parameter $v \in [0,1]$ is the weight for the controlled slack variable; Φ is a map from the non-empty set of the original input data. The decision function is defined as in (6):

$$f(x) = sgn(\sum_{i=1}^{n} \alpha_i k(x_i x) - \rho) \tag{6}$$

where x represents the input data vector, and k represents kernel function which is used for parameter selection in SVM. This is defined as in (7):

$$k(x, x') = \langle \Phi(x) \cdot \Phi(x') \rangle \tag{7}$$

3.4.2. Decision tree

Typically using a top-down greedy method, a DT classifier offers a quick and efficient way to classify data instances [23]. Training datasets are recursively divided into smaller subsets by DT until every set is part of a single class. Information theory is used iteratively by the DT algorithm as a means of selecting attributes. In every sample space, there are data that might not contribute to the decision made by the DT model; the model tries to make its decision by ensuring that the decision boundaries are void of impurities as much as possible. Entropy computation has been formalized by Shannon. Let assume a random variable X with values x_i and probability pr(x_i) has its entropy in (8):

$$H(X) = -\sum_i \Pr(x_i) log_2 \Pr(x_i) \quad (8)$$

In addition, information gain, which is meant to be maximized in the decision processes, has the lowest value as zero (0) and the highest value as one (1). In some other texts, this is called gain ratio, which draws many of relationships from the entropy. It is mostly defined by the difference between the initial and the final entropy, as shown in (9):

$$IG(X,i) = H(X) - H(X|i) \quad (9)$$

After applying each classification (C) algorithm to the selected data feature (x) separately, this research computes the result of each of the model classification results from each test instance, and the final output is predicted using a voting mechanism as computed in (10):

$$y = Max\{C_1(x), C_2(x), \ldots., C_n(x)\} \quad (10)$$

where x is the extracted input data and C is the assigned classifier.

3.5. Hyperparameter tuning

To obtain the experimental results and achieve the objective of this study all algorithms were coded in python language. The data obtained were separated as training and test data to test the algorithm that can give optimum stock price prediction performance. The hyperparameters used in the implementation are shown in Table 1.

Table 1. Hyperparameter used for the implementation

Model	Parameter
SVM	C: [0.1, 0.5, 1, 1.5]
	Epsilon: [0.01, 0.1, 1]
	Gamma: [‘auto’]
	Kernel: [‘linear’, ‘poly’,’rbf’]
RF	Max_depth: [80, 90, 100, 110]
	Max-features: [2, 3]
	Min-samples-leaf: [3, 4, 5]
	Min_samples_split: [8, 10, 12]
	No_estimator: [100, 200, 300,1000]
DT	Max_feature: [5, 210,13, 20]
	Max_sample: [50, 200, 400, 600]
	Max_depth: [2, 4, 6, 8, 10]
	Criterion” [‘gini’, ‘entropy’]

For the optimization of the algorithms, it is crucial to find the optimal parameters instead of directly giving values. In this study, to find the optimum parameters, the parameter with the highest accuracy were selected from the hyperparameter in Table 1 using grid search approach.

3.6. Modelling training and testing

In this stage, the different classifiers are trained on the stock market dataset. A cross-validation technique is used to test the model using the validation dataset after it has been trained on the annotated dataset, with 90% of the dataset designated for testing and the remaining 10% for training. This is because training and testing datasets need to be appropriate representations of stock market price. Overfitting and bias were prevented by repeating this procedure. The best machine-learning techniques were determined by testing the trained model on stock dataset.

3.6.1. Algorithmic and mathematical analysis

Algorithm 1 displays the pseudo-code used to implement LR with SVM model. After the features have been extracted from the stock market dataset, the classifier is applied to predicted the target value which is the stock price.

Algorithm 1. Prediction of stock market price using hybrid LR and SVM

Input: Training dataset {xi,yi}i=1N, text sampel x, k(1≤k≤N),p(p≥1)
Output: Prediction of stock price yˆfor xˆ

```
for i =1:k
  for each training data instance d_i
    set the tarhet value for the regression to Z_i
    initialize the weith of instance d_j to [P(1|d_j(1-P(1|d_j))]
    finalize f(j) to the data with class value (Z_i) and weight (w_j)
    assign class label 1
    if P_id > 0.5 otherwise
      class label 2
    endif
  endfor
endfor
//SVM
Normalize the dataset
  for each C, γ
    class validation using leave-one-out
    train and test SVM
    store the success rate
endfor
 compute the average success rate
update the best C, γ
choose C, γ with the best success rate
return ŷ
```

3.7. Evaluation metrics

To evaluate the performance of the new model, the cross-validation technique is adopted [27]. The cross-validation technique is one of the scoring evaluation schemes used to measure the performance of the classifier [28]. This technique involves splitting the dataset into 90% for training and 10% for testing. Furthermore, the performance of different stock prediction models is evaluated based on the following metrics: mean absolute error (MAE), mean squared error (MSE), mean absolute percentage error (MAPE), RMSE, and R^2 [29]. These are computed as in (12)–(16) respectively.

MAE: this is the average of the absolute different between the predicted and actual values as shown in (12):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \tag{12}$$

where x_i and y_i represent the actual and predicted values, and n is the total number of stock samples.

MSE: this is defined as the average squared differences between the actual value and predicted values, as shown in (13):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 \tag{13}$$

where x_i and y_i represent the actual and predicted values, and n is the total number of stock samples.

MAPE: this is defined as the average absolute percentage difference between predicted values and actual values as in (14):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - y_i}{y_i}\right| \tag{14}$$

where x_i and y_i represent the actual and predicted values, and n is the total number of stock samples.

RMSE: it is a predictive model by calculating the square root of the average of squared differences between prediction and actual observation, as in (15):

$$RMSE = \sqrt{\frac{1}{n}\sum_{n}^{n}(x_i - y_i)^2} \tag{15}$$

A smaller RMSE signifies a better fit of the model to the data, indicating the model's predictions are closely aligned with the actual values. Where x_i and y_i represent the actual and predicted values, and n is the total number of stock samples.

R^2: also recognized as the coefficient of determination, is a statistical index that quantifies the extent to which changes in the dependent variable can be attributed to its relationship with one or more independent variables as in (16):

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=0}^{n}(y_i - \widehat{y_i})^2} \tag{16}$$

where n is the total number of samples, x_i and y_i represent the actual and predicted values.

The closer the value of 1, the better the value of R^2. The lowest value of MSE, MAE, RMSE, and MAPE is the best value [30]. The next section discusses the experimental evaluations and result.

4. EXPERIMENTAL EVALUATIONS AND RESULTS

This section presents different experiments carried out to achieve the prediction of stock market price using the proposed method. The configuration and parameter settings and the experimental implementations are discussed in the subsequent part.

4.1. Data description and experimental settings

The implementation software used in this research is Python software. The training and testing procedures were conducted on a computer equipped with a both central processing unit (CPU) and graphical processing unit (GPU) computers. The stock dataset is obtained from the Yahoo Finance stock market and Dhaka stock market database repositories [26], [31].

- Yahoo Finance stock market dataset

The Yahoo Finance stock market dataset used in this implementation starts from April 1, 2018 to March 31, 2023, and contains 1,257 rows and 7 columns [26]. The dataset is a financial dataset containing daily stock market data for multiple assets. The features such as the open, high, low, close, adj_close which are extracted to improve the prediction performances of the classifiers. These features description, sample of yahoo and Dhaka finance stock datasets are further explained in Tables 2-4.

Table 2. Features description

Features	Descriptions
Date	This is the date on which the stock market data was recorded
Open	This is the opening price of the asset on the given date
High	This is the highest price of the asset on the given date
Low	This is the lowest price of the asset on the given date
Close	This is the closing price of the asset on the given date
Adj close	This is the adjusted closing price of the asset on the given date. This price takes into account any dividends, stock splits, or other corporate actions that may have occurred, which can affect the stock price
Volume	This is the total number of shares of the asset that were traded on the given date
Date	his is the date on which the stock market data was recorded

Table 3. Sample of Yahoo Finance stock data

Date	Open	High	Low	Close	Adj close	Volume
28-04	33,797.43	34,104.56	33,728.40	34,098.16	34,098.16	354,310
27-04	33,381.66	33,859.75	33,374.65	33,826.16	33,826.16	343,240
26-04	33,596.34	33,645.83	33,235.85	33,301.87	33,301.87	321,170
06-02	33,874.44	33,962.84	33,683.58	33,891.02	33,891.02	296,400
28-11	34,275.91	34,303.88	33,799.35	33,849.46	33,849.46	300,330
25-11	34,213.04	34,386.51	34,199.57	34,347.03	34,347.03	131,660
23-11	34,091.57	34,262.07	34,004.64	34,194.06	34,194.06	236,820
17-11	33,329.27	33,616.02	33,239.75	33,546.32	33,546.32	319,210

To avoid biases in the implementation and for proposed model generalization, another a publicly available Dhaka stock market dataset was used [31]. The dataset starts from January 2013 to February 2024. The samples of the dataset as shown in Table 4.

Table 4. Sample of Dhaka stock market data

Date	Open	High	Low	Close	Volume
28-01	1460.3	1476.26	1459.67	1473.01	833692000
29-01	1473.01	1500.77	1473.01	1495.67	1326800000
30-01	1495.67	1506.46	1485.94	1486.56	1054900000
31-01	1486.56	1495.53	1483.61	1488.56	925494000
03-02	1488.56	1499.44	1481.38	1494.15	1081130000
03-04	1494.15	1513.3	1494.15	1503.97	1330170000
02-05	1503.97	1521.76	1498.16	1518.42	1015790000

4.2. Experimental results and analysis

In this section, the results of the proposed algorithm on the publicly available Yahoo Finance stock market and Dhaka stock market databases [26] are presented to assess the performance of the proposed method compared with other baseline detection methods.

4.3. Experiment 1: qualitative evaluation of the proposed model and the baseline methods on Yahoo Finance stock dataset

This study aims to the performance of the proposed method with other baseline ML techniques used in this study on yahoo finance stock market for the prediction of stock prices. The actual and predicted data is shown in Table 5.

Table 5. Quantitative evaluation of proposed model with baseline predictive models on Yahoo Finance stock market dataset

	Actual growth rate value	Random forest	Decision tree	Proposed method
	Actual values	Predicted values	Predicted values	Predicted values
0	0.434694	0.431725	0.431725	0.431725
1	0.431216	0.427996	0.427996	0.427996
2	0.429354	0.429508	0.429508	0.429508
3	0.430333	0.414721	0.414721	0.414721
4	0.416166	0.415496	0.415496	0.415496
5	0.417105	0.415134	0.415134	0.415134
6	0.413863	0.425594	0.425594	0.425594

Table 5 shows the actual and predicted stock market prices using the DT, RF, and proposed method. When compared to other implementation methods, proposed method produced a better prediction which is as close to the actual prediction as it can be seen in Table 5. Additionally, the performance of each method is evaluated by quantitative evaluation in terms of the MSE, MAE, RSME, MAPE, R^2 score, and accuracy. The summarized results of other metrics used in the implementation is shown in Table 6.

Table 6. Performance evaluation of ML baseline methods (Yahoo Finance stock dataset)

Model	MSE	MAE	RSME	MAPE	R^2	Accuracy
RF	0.02051	0.0347	0.45288	0.3759	0.76	93.91
DT	0.01518	0.0235	0.38969	0.1429	0.94	95.43
Proposed model	0.00805	0.0180	0.29674	0.1073	0.99	97.15

Table 6 presents an additional analysis of the stock prices generated performance to MSE, MAE, RSME, MAPE, R^2 score, and accuracy. Based on the outcome displayed in Table 6, the proposed method has an MSE 0.00805, MAE of 0.0180, RSME of 0.29674, MAPE of 0.1073, R^2 score of 0.99, and accuracy of 97.15. The MSE value of 0.00805 suggests that, on average, the model's forecast exhibit a deviation approximately 0.0008 unit. The RMSE of 0.29674 signifies that the average deviation between predicted and actual values is merely 0.3 unit. The MAE score of 0.0180 signifies that the proposed model's prediction exhibits an average variance of merely 0.018 units from actual data. The R^2 of 0.99 suggests that the proposed method exhibits a high degree of accuracy in capturing the stock price patterns present in the data and effectively accounts for a significant percentage of variation in the independent variables. The stronger R^2 value suggest that the model effectively account for the variability of the observed data.

4.4. Experiment 2: qualitative evaluation of the proposed model and the baseline methods on Dhaka stock market dataset

This study aims to generalize the performance of the proposed method and other baseline ML techniques used in this study on Dhaka stock market for the prediction of stock prices. The actual and predicted data is shown in Table 7.

Table 7. Quantitative evaluation of proposed model with baseline predictive models on Dhaka stock market dataset

	Actual growth rate value Actual values	Random forest Predicted values	Decision tree Predicted values	Proposed method Predicted values
0	0.455200	0.438992	0.431725	0.438992
1	0.438566	0.433367	0.427996	0.433367
2	0.432565	0.438992	0.429508	0.438992
3	0.440565	0.433367	0.414721	0.433367
4	0.431265	0.433367	0.415496	0.433367
5	0.434539	0.416492	0.415134	0.416492
6	0.414476	0.422117	0.425594	0.422117

Table 7 shows the actual and predicted stock market prices using the DT, RF, and proposed method. When compared to other implementation methods, proposed method produced a better prediction which is as close to the actual prediction as it can be seen in Table 7. Furthermore, the performance of each method is evaluated by quantitative evaluation in terms of the MSE, MAE, RSME, MAPE, R^2 score, and accuracy. The summarized results of other metrics used in the implementation is shown in Table 8.

Table 8 presents an additional analysis of the stock prices generated performance to MSE, MAE, RSME, MAPE, R^2 score, and accuracy. Based on the outcome displayed in Table 8, the proposed method has an MSE 0.00805, MAE of 0.0044, RSME of 0.0085, MAPE of 0.1002, R^2 score of 0.9835, and accuracy of 98.86. The MSE value of 0.0001 suggests that, on average, the model's forecast exhibit a deviation approximately 0.0008 unit. The RMSE of 0.0085 signifies that the average deviation between predicted and actual values is merely 0.009 unit. The MAE score of 0.0044 signifies that the proposed model's prediction exhibits an average variance of merely 0.004 units from actual data. The R^2 of 0.9835 suggests that the proposed method exhibits a high degree of accuracy in capturing the stock price patterns present in the data and effectively accounts for a significant percentage of variation in the independent variables. The stronger R^2 value suggest that the model effectively account for the variability of the observed data.

Table 8. Performance evaluation of ML baseline methods (Dhaka stock market dataset)

Model	MSE	MAE	RSME	MAPE	R^2	Accuracy (%)
RF	0.0013	0.0356	0.0364	0.3759	0.7004	70.11
DT	0.0002	0.0155	0.0111	0.9909	0.9450	93.82
Proposed model	0.0001	0.0044	0.0085	0.1002	0.9835	98.86

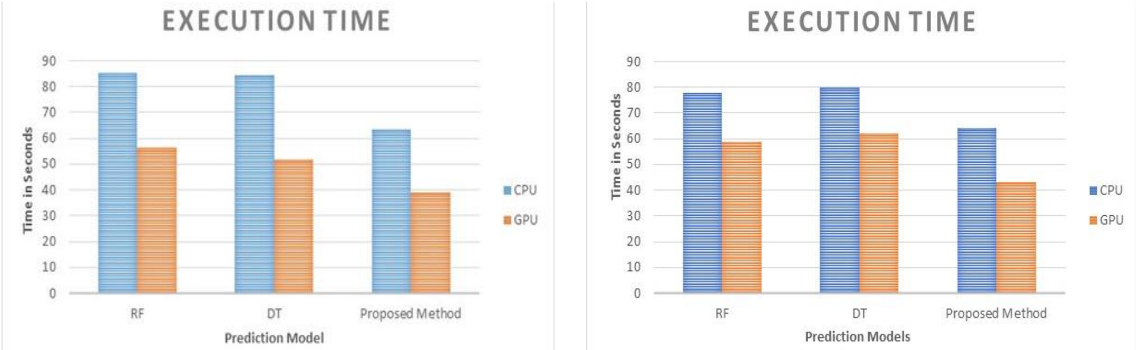
4.5. Comparison evaluation of the proposed model with execution time

The execution time of the models used in the implementation on the Yahoo Finance stock and Dhaka stock datasets for the prediction of stock prices on both CPU and GPU are as shown in Figure 2. From Figures 2(a) and (b), one can be observed that the execution time of the proposed model is faster than RF and DT models used in the implementation on CPU, and this is also applicable to GPU on both Yahoo Finance stock and Dhaka stock datasets, thereby addressing a critical aspect of real-time stock prediction.

4.6. Comparison evaluation of performance of the proposed method with other methods

This section compares the performance of the proposed model with other state-of-the-art predictive models that are currently in use on Yahoo Finance stock datasets regarding the following: MAE, RSME, R^2 score, and accuracy, and model used, as shown in Table 9.

Table 9 shows that performance of the proposed model with other models in literature on Yahoo Finance stock and Dhaka stock market datasets. The proposed method on Yahoo Finance dataset has MAE of 0.0180, the RMSE score is 0.29674, the R^2 is 0.990 compared to another model in the Table 9. This attests to the inference made in [37] that the closer the value of 1, the better the value of R^2. The lowest value of MSE, MAE, RMSE, and MAPE is the best value. The proposed model achieved accuracy of 97.15% which were higher than other methods in Table 9. Furthermore, on the Dhaka stock market dataset, the proposed method shows superior performance in comparison with other models used in the literature. The proposed method, however, yields excellent prediction results based on the six-evaluation metrics used in the implementation.



(a) (b)

Figure 2. Comparative analysis of execution time of; (a) each model on CPU and GPU on Yahoo Finance stock dataset and (b) each model on CPU and GPU on Dhaka stock dataset

Table 9. Performance evaluation of proposed model with other existing prediction models

Authors	Model	Dataset used	MAE	RSME	R2-score	Accuracy (%)
[32]	NMC, BERT, LSTM, and DQN	Dhaka	0.104	0.0739		61.77
[33]	ARIMA	Dhaka	-	0.8705	0.9259	-
[34]	Ensemble of graph theory	Dhaka	5-10	10-15	-	-
[35]	Improved KNN	Dhaka	0.394	0.394	0.988	-
[36]	XGBoost	Dhaka	0.558	-	0.989	-
Proposed method	LR and SVM	Yahoo Finance	0.0180	0.2967	0.990	97.15
		Dhaka	0.0044	0.0085	0.9835	98.86

5. DISCUSSION

This study investigated the prediction of stock market prices using the integration of LR and SVM on publicly available Yahoo Finance stock and Dhaka stock market datasets. From the experiments conducted, we found that the LR provides interpretability as it estimates the probabilities, making it easier to understand the impact of the features on the stock market price movements, the SVM handles the nonlinear relationships between features and target variables, potentially capture more complex patterns in the data, as shown in Tables 5 and 7. Furthermore, the execution time of the proposed method was 64 seconds on CPU and 38 seconds on GPU and 66 seconds on CPU and 43 seconds on GPU on Yahoo Finance and Dhaka finance datasets respectively as shown in Figure 2. The proposed method used in this study tended to have a higher proportion of prediction accuracy of 97.15% on Yahoo Finance dataset and accuracy of 98.86% on Dhaka stock market dataset. In comparison of the proposed method with other prediction methods used in literature as shown in Table 9, the study suggests that higher accuracy. The proposed method may benefit from the strengths of two models' integration.

Limitations: this study explored comprehensive and advanced approaches to deal with the issue of predicting stock market prices for potential investors in addition to accurate stock price prediction hampered by the computing efficiency of real-time stock market. However, further, and in-depth studies may be needed to confirm the computational complexity and memory-intensive especially in large datasets, making them less scalable for extensive feature sets or long time series. Also, the model sensitivity to outliers which are common in financial data can affect the model's prediction performances.

6. CONCLUSION

As the stock market is bound tightly to a country's economic growth and brings in huge investments by the investors and issues equities in the public interest, forecasting the movement of the stock prices and the market becomes essential to prevent huge losses and make relevant decisions.

This study has presented a novel framework that integrates LR with SVM for enhanced stock market price prediction. Combination of the interpretability of LR with the robustness of SVM achieved superior predictive performance and facilitates informed investment decision-making. This research highlights the efficacy of integrating LR and SVM into a hybrid model to improve prediction accuracy in complex classification tasks. The key findings demonstrate that by combining the linear capabilities of LR with the non-linear flexibility of SVM, the hybrid model effectively captures a broader range of data patterns. This integration leverages the strengths of both methods, balancing bias and variance, and enhancing the model's ability to generalize across diverse datasets. The research contributes a robust methodology for model integration, including strategies such as votingwhich collectively lead to significant improvements in predictive performance. This hybrid approach not only offers superior accuracy but also provides a versatile framework for tackling real-world challenges where data complexity and variability are prevalent.

This study demonstrates that utilizing the proposed method is more resilient than the other state-of-the-art methods used in this study. Future studies may explore the use of more temporal features with feasible ways of establishing influence of period, holiday and time variables to resolve the inaccurate decision making by investors in prediction of stocks prices. Future research directions include exploring additional hybrid modeling techniques and incorporating alternative data sources to further improve the accuracy and reliability of stock market prediction models.

ACKNOWLEDGMENTS

Author thanks the Department of Network and Support, and Department of Business Applications and Development, Walter Sisulu University for the resources and financial support made available.

FUNDING INFORMATION

There is state no funding involved.

AUTHORS CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Omobayo Ayokunle Esan			✓	✓	✓				✓	✓	✓	✓	✓	✓
Dorcas Oladayo Esan	✓	✓				✓		✓		✓	✓			
Femi Abiodun Elegbeleye		✓			✓	✓	✓	✓		✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

ETHICAL APPROVAL

The research does not require ethical approval since the dataset used for the implementation does not involves human or animal, and the datasets are publicly available.

DATA AVAILABILITY

The data that support the findings of this study are openly available in kaggle at https://www.kaggle.com/datasets/suruchiarora/yahoo-finance-dataset-2018-2023, and Dhaka stock dataset in https://www.worldscientific.com/doi/full/10.1142/S146902682350013X.

REFERENCES

[1] M. Mabeba, "The Effect of Financial Market Depth on Economic Growth in Developing Countries with Large Financial Sectors," *Social Science Studies*, vol. 4 no. 2, pp. 66-81, 2024, doi: 10.47153/sss42.8022024.

[2] B. S. Kureti and N. S. Mothe, "Stock Trend Prediction Using KNN," *International Journal of Advances in Engineering and Management (IJAEM)*, vol. 5, no. 4, pp. 277-279, 2023, doi: 10.35629/5252-0504277279.

[3] N. Rouf *et al.*, "Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions," *Electronics*, vol. 10, no. 2717, pp. 1-25, 2021, doi: 10.3390/electronics10212717.

[4] J. Zou *et al.*, "Stock Market Prediction via Deep Learning Techniques: A Survey," *arXiv*, 2023, doi: 10.48550/arXiv.2212.12717.

[5] K.-Y. Woo, C. Mai, M. McAleer, and W.-K. Wong, "Review on Efficiency and Anomalies in Stock Markets," *Economies,* vol. 8, no. 1, pp. 1-51, 2020, doi: 10.3390/economies8010020.

[6] I. Setiani, M. N. Tentua, and S. Oyama, "Prediction of Banking Stock Prices Using Naïve Bayes Method," *Journal of Physics: Conference Series* no. 1823 pp. 1-7, 2021, doi: 10.1088/1742-6596/1823/1/012059.

[7] M. Li, Y. Zhu, Y. Shen, and M. Angelova, "Clustering-enhanced stock price prediction using deep learning," *World Wide Web*, vol. 26, pp. 207–232, 2023, doi: 10.1007/s11280-021-01003-0.

[8] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, E. Salwana, and S. Shahab, "Deep Learning for Stock Market Prediction," *Entropy*, vol. 22, no. 8, pp. 1-23, 2020, doi: 10.3390/e22080840.

[9] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "Efficient Stock-Market Prediction Using Ensemble Support Vector Machine," *Open Computer Science,* vol. 10, no. 1, pp. 154-163, 2020, doi: 10.1515/comp-2020-0199.

[10] G. Sonkavde, D. S. Dharrao, A. M. Bongale, S. T., D. D. Deokate, and S. K. Bhat, "Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications," *International Journal of Financial Studies* vol. 11, no. 94, pp. 1-22, 2023, doi: 10.3390/ijfs11030094.

[11] S. Mukherjee, B. Sadhukhan, N. Sarkar, D. Roy, and S. De, "Stock market prediction using deep learning algorithms," *CAAI Transactions on Intelligence Technology,* vol. 8, no. 2023, pp. 82–94, 2023, doi: 10.1049/cit2.12059.

[12] I. Akin and M. Akin, "Behavioral finance impacts on US stock market volatility: an analysis of market anomalies," *Behavioural Public Policy* pp. 1-25, 2024, doi: 10.1017/bpp.2024.13.

[13] G. Sonkavde, D. S. Dharrao, A. M. Bongale, S. T. Deokate, D. Doreswamy, and S. K. Bhat, "Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications," *International Journal of Financial Studies*, vol. 11, no. 3, pp. 1-22, 2023, doi: 10.3390/ijfs11030094.

[14] C. Y. Lin and J. A. L. Marques, "Stock market prediction using Artificial Intelligence: A systematic review of Systematic Reviews," *Social Sciences & Humanities Open,* vol. 9, no. 100864, pp. 1-11, 2024, doi: 10.1016/j.ssaho.2024.100864.

[15] D. Selvamuthu, V. Kumar, and A. Mishra, "Indian stock market prediction using artificial neural networks on tick data," *Financial Innovation,* vol. 5, no. 16, 2019, doi: 10.1186/s40854-019-0131-7.

[16] J. M. Sangeetha and K. J. Alfia, "Financial stock market forecast using evaluated linear regression based machine learning technique," *Measurement: Sensors,* vol. 31 no. 100950, pp. 1-7, 2024, doi: 10.1016/j.measen.2023.100950.

[17] D. Dave, S. Prajapati, and U. Solanki, "Stock Price Prediction using Machine Learning," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 9, pp. 263-275, Sep. 2022, doi: 10.22214/ijraset.2022.46610.

[18] M. Naeem, H. S. Jassim, and D. Korsah, "The Application of Machine Learning Techniques to Predict Stock Market Crises in Africa," *Journal of Risk and Financial Management,* vol. 17, no. 12, pp. 1-19, 2024, doi: 10.3390/jrfm17120554.

[19] S. L. V. V. D. Sarma, D. V. Sekhar, and G. Murali, "Stock market analysis with the usage of machine learning and deep learning algorithms," *Bulletin of Electrical Engineering and Informatics,* vol. 12, no. 1, pp. 552-560, 2023, doi: 10.11591/eei.v12i1.4305.

[20] S. Sang and L. Li, "A Novel Variant of LSTM Stock Prediction Method Incorporating Attention Mechanism," *Mathematics*, vol. 12, no. 7, pp. 1-20, 2024, doi: 10.3390/math12070945.

[21] T. Julian, T. Devrison, V. Anora, and K. M. Suryaningrum, "Stock Price Prediction Model Using Deep Learning Optimization Based on Technical Analysis Indicators," *Procedia Computer Science,* vol. 227, pp. 939-947, 2023, doi: 10.1016/j.procs.2023.10.601.

[22] D. Ahmed, R. Neema, N. Viswanadha, and R. Selvanambi, "Analysis and Prediction of Healthcare Sector Stock Price Using Machine Learning Techniques: Healthcare Stock Analysis," *International Journal of Information System Modeling and Design (IJISMD),* vol. 13, no. 9, pp. 1-15, 2017, doi: 10.4018/IJISMD.303131.

[23] K. Chahuán-Jiménez, "Neural Network-Based Predictive Models for Stock Market Index Forecasting," *Journal of Risk and Financial Management,* vol. 17, no. 6, pp. 1-18, 2024, doi: 10.3390/jrfm17060242.

[24] U. R. Saxena, P. Sharma, and G. Gupta, "Comprehensive Study of Machine learning Algorithms for Stock Market Prediction during COVID-19," *Journal of Computers, Mechanical and Management,* vol. 1, no. 2, pp. 14-21, 2022, doi: 10.57159/gadl.jcmm.1.2.22021.

[25] J. Shen and M. O. Shafiq, "Short-term stock market price trend prediction using a comprehensive deep learning system," *Journal of Big Data,* vol. 7, no. 66, pp. 1-33, 2020, doi: 10.1186/s40537-020-00333-6.

[26] Y. Finance, "Yahoo Finance Stock Market Prediction," [Online]. Available: https://www.kaggle.com/code/hamidsadeghi1/yahoo-finance-stock-market-prediction/notebook, 2024.

[27] O. A. Esan and I. O. Osunmakinde, "Crime Prediction Linked to Geographical Locations with Periodic Features for Societal Security," *2023 15th International Conference on Computer Research and Development (ICCRD),* pp. 6-14, 2023, doi: 10.1109/ICCRD56364.2023.10080235.

[28] D. Abdelhafiz, C. Yang, R. Ammar, and S. Nabavi, "Deep convolutional neural networks for mammography: advances, challenges and applications," *7th IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS 2017),* vol. 20, pp. 76-103, doi: 10.1186/s12859-019-2823-4, 2017.

[29] D. O. Esan, P. A. Owolawi, and C. Tu, "Anomalous Detection in Noisy Image Frames using Cooperative Median Filtering and KNN," *IAENG International Journal of Computer Science,* vol. 49, no. 1, pp. 1-10, 2022.

[30] M. K. Ho, H. Darman, and S. Musa, "Stock Price Prediction Using ARIMA, Neural Network and LSTM Models," *Journal of Physics: Conference Series,* vol. 1988, no. 012041, pp. 1-12, 2021, doi: 10.1088/1742-6596/1988/1/012041.

[31] T. Muhammad, R. Islam, and M. S. Alam, "Dhaka Stock Exchange Historical Data," *Mendeley Data,* vol. 3, 2024, doi: 10.17632/23553sm4tn.3.

[32] C. Liu, J. Yan, F. Guo, and M. Guo, "Forecasting the Market with Machine Learning Algorithms: An Application of NMC-BERT-LSTM-DQN-X Algorithms in Quantitative Testing," *ACM Transactions on Knowledge Discovery from Data (TKDD),* vol. 16, no. 4, pp. 1-22, 2022, doi: 10.1145/348837.

[33] M. I. Omar and M. Rahaman, "Stock Price Trend Forecasting using Long Short Term Memory Recurrent Neural Networks," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology,* vol. 6, no. 4, pp. 468-474, 2020, doi: 10.32628/CSEIT206474.

[34] P. Patil, C. S. M. Wu, K. Potika, and M. Orange, "Stock Market Prediction using an Ensemble of Graph Theory, Machine Learning and Deep Learning Models," *Conference: ICSIM '20: The 3rd International Conference on Software Engineering and Information Management*, pp. 85-92, 2020, doi: 10.1145/3378936.3378972.

[35] Y. Wang, Y. Xie, Y. Wu, and Y. Yang, "Improved KNN-based Stock Price Prediction," *Academic Journal of Computing & Information Science*, vol. 7, no. 6, pp. 38-43, 2024, doi: 10.25236/AJCIS.2024.070606.

[36] D. Kocaoglu, K. Turgut, and M. Z. Konyar, "Sector-Based Stock Price Prediction with Machine Learning Models," *Sakarya University Journal of Computer and Information Sciences*, vol. 5, no. 3, pp. 415-426, 2022, doi: 10.35377/saucis.05.03.1200151.

[37] J. K. Lubis and I. Kharisudin, "Optimal Modeling of Stock Price Forecasting using Long Short-Term Memory and Auto ARIMA Methods," *UNNES Journal of Mathematics*, vol. 11, no. 2, pp. 141-152, 2022.

BIOGRAPHIES OF AUTHORS

Omobayo Ayokunle Esan has a Ph.D. in Computer Science-School of Computing, Faculty of Science, Engineering and Technology, University of South Africa, South Africa. He is also a lecturer at Department of Networking and Support in Walter Sisulu University, Postsdam Campus, East London, South Africa. He can be contacted at email: oesan@wsu.ac.za.

Dorcas Oladayo Esan has a doctoral degree in Computer Systems Engineering, Faculty of Information Communication Technology (ICT), at Tshwane University of Technology, South Africa. She can be contacted at email: oladayojadesola10@gmail.com.

Femi Abiodun Elegbeleye is a doctoral student at Northwest Univerisy, South Africa in the Department of Computer Science. He is also a lecturer at Department of Business Application Development in Walter Sisulu University, Postsdam Campus, East London, South Africa. He can be contacted at email: felegbeleye@wsu.ac.za.